Sarang Kashyap
sgkashya
Path 2: Bike Traffic
https://github.com/ECEDataScience/miniproject-f23-sgkashya

Dataset Description:

The dataset I chose to work with was the bike usage in New York City. This dataset gave information on bike traffic across a number of bridges in New York City, specifically the Brooklyn, Manhattan, Williamsburg, and Queensboro Bridges. The .csv file given for this path describes the parameters for this dataset, which include high/low temperatures, precipitation, bicycle counts for each bridge, and the total bike usage in each day.

Method:

1) Sensor Installation on Bridges:
   - For this question, I chose to conduct a correlation analysis because this would be the most efficient way to find which of the four bridges should have sensors installed on to get the best prediction of overall traffic. By using this method, I was able to find the correlation coefficients between each bridge's traffic and the total traffic from all bridges combined. From this calculation, I was expecting that the higher correlation coefficients would indicate which bridges had a stronger correlation with total traffic, and so I would be able to choose the highest three coefficients to provide the best overall traffic estimate.

2) Predicting Traffic Based on Weather Forecast:
   - For this question, I also chose to conduct correlation analysis because this would be the most efficient way to predict the total number of bicyclists for each day from the day's weather forecast. By using this method, I was able to find the correlation coefficients between the total bicycle traffic and each weather factor (high/low temperature and precipitation). From this calculation, I was expecting that the higher correlation

coefficients would indicate which weather factor had a stronger correlation with the total number of bicyclists, and so I would be able to predict the total number of bicyclists for each day.

3) Predicting the Day of the Week Based on Bicycle Counts:
- For this final question, I chose to conduct an analysis using a Random Forest Classifier. The expectation of using this method was to predict the day of the week by obtaining a categorical variable for model accuracy based on the number of bicyclists on the bridges. I was hoping that this variable for model accuracy would tell me how accurate the prediction of the day would be, so a lower number would mean that the number of bicyclists is not a great indication of what day of the week it is.

Results:

1) Sensor Installation on Bridges:
- The results of the correlation analysis I performed for this question showed that the three bridges with the highest correlation coefficients were the Williamsburg Bridge (0.975), Queensboro Bridge (0.963), and Manhattan Bridge (0.935). These are the three bridges that should have sensors installed on to get the best prediction of overall traffic.

2) Predicting Traffic Based on Weather Forecast:
- The results of the correlation analysis I performed for this question showed that weather forecasts with higher temperature would generate the most number of bicyclists in a day. The correlation coefficient for high temperature was 0.574, whereas for low temperature it was 0.442, and for precipitation it was -0.421. To answer the question: yes, the total number of bicyclists in a day can be predicted by the weather forecast for that day.

3) Predicting the Day of the Week Based on Bicycle Counts:
- The results of the analysis using the Random Forest Classifier show that the number of bicyclists are not a strong predictor of the day of the week. I came to this conclusion from the model

accuracy of 24.6%, which means that the model predicts the correct day based on the number of bicyclists only 24.6% of the time, thus proving that there is no strong prediction for the day of the week based on the number of bicyclists. To answer the question: we can not use the data to accurately predict what day it is today based on the number of bicyclists on the bridges.

Analysis:

1) Sensor Installation on Bridges:
   - The results I got for this question mean that the Williamsburg Bridge, Queensboro Bridge, and Manhattan Bridge would best represent the overall traffic pattern because they had the three highest correlation coefficients, whereas the Brooklyn Bridge had the lowest correlation with a coefficient of 0.874.

2) Predicting Traffic Based on Weather Forecast:
   - The results I got for this question mean that there is a positive correlation for high temperature and low temperature with bicycle traffic, with higher temperature has the stronger correlation. On the other hand, since precipitation has a negative correlation, this means that it is associated with decreased bicycle traffic on a certain day. Therefore, warmer days have the most impact on bicycle traffic, whereas rainy days have the least impact on bicycle traffic.

3) Predicting the Day of the Week Based on Bicycle Counts:
   - The results I got from this question mean the level of accuracy is pretty low for predicting the day of the week, if based on only the number of bicyclists. This could be because there are other factors that impact the number of bicyclists in a day, such as the weather forecast or other factors that are not in our knowledge.